

PREDIKCIJA KIŠE

Nenad Raković SW45/2019
Marko Dragičević SW74/2019

UVOD

Problem koji se rešava jeste predikcija kiše za sutrašnji dan u odnosu na dnevne podatke u proteklih deset godina. Cilj projekta je analizirati koji od testiranih algoritama ima najbolje rezultate. Koristili smo: **logističku regresiju**, **XGBoost** i **ANN**.

PODACI

Početni skup podataka sadrži:

- **145.460** redova
- **23** obeležja



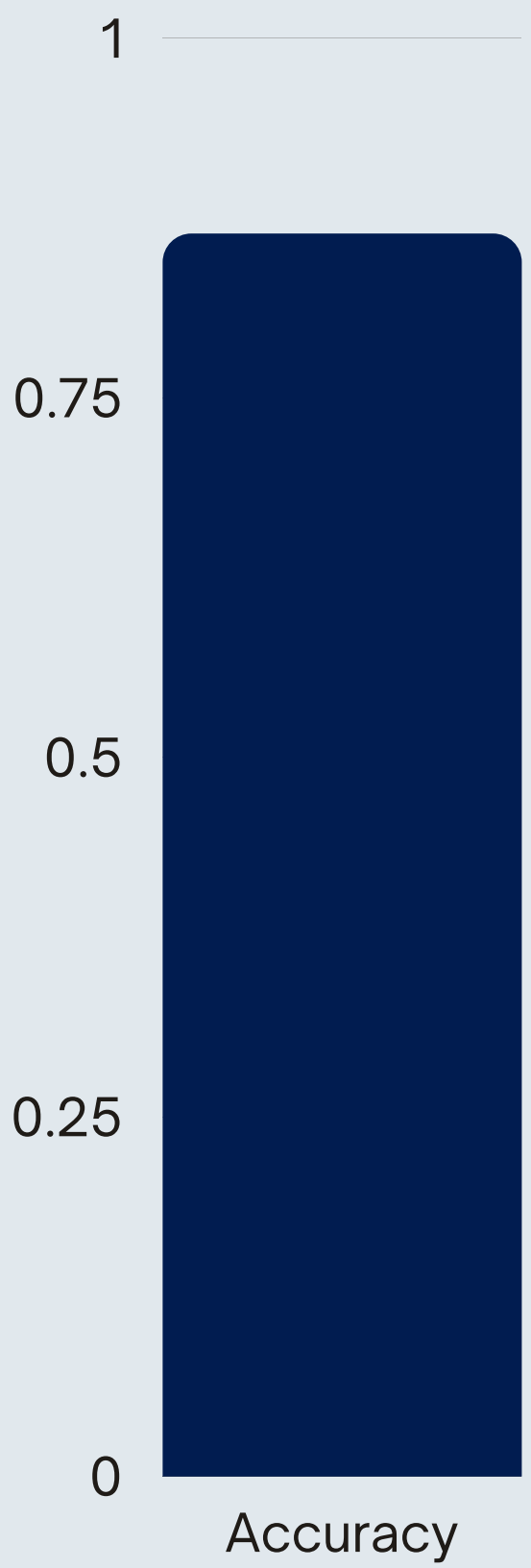
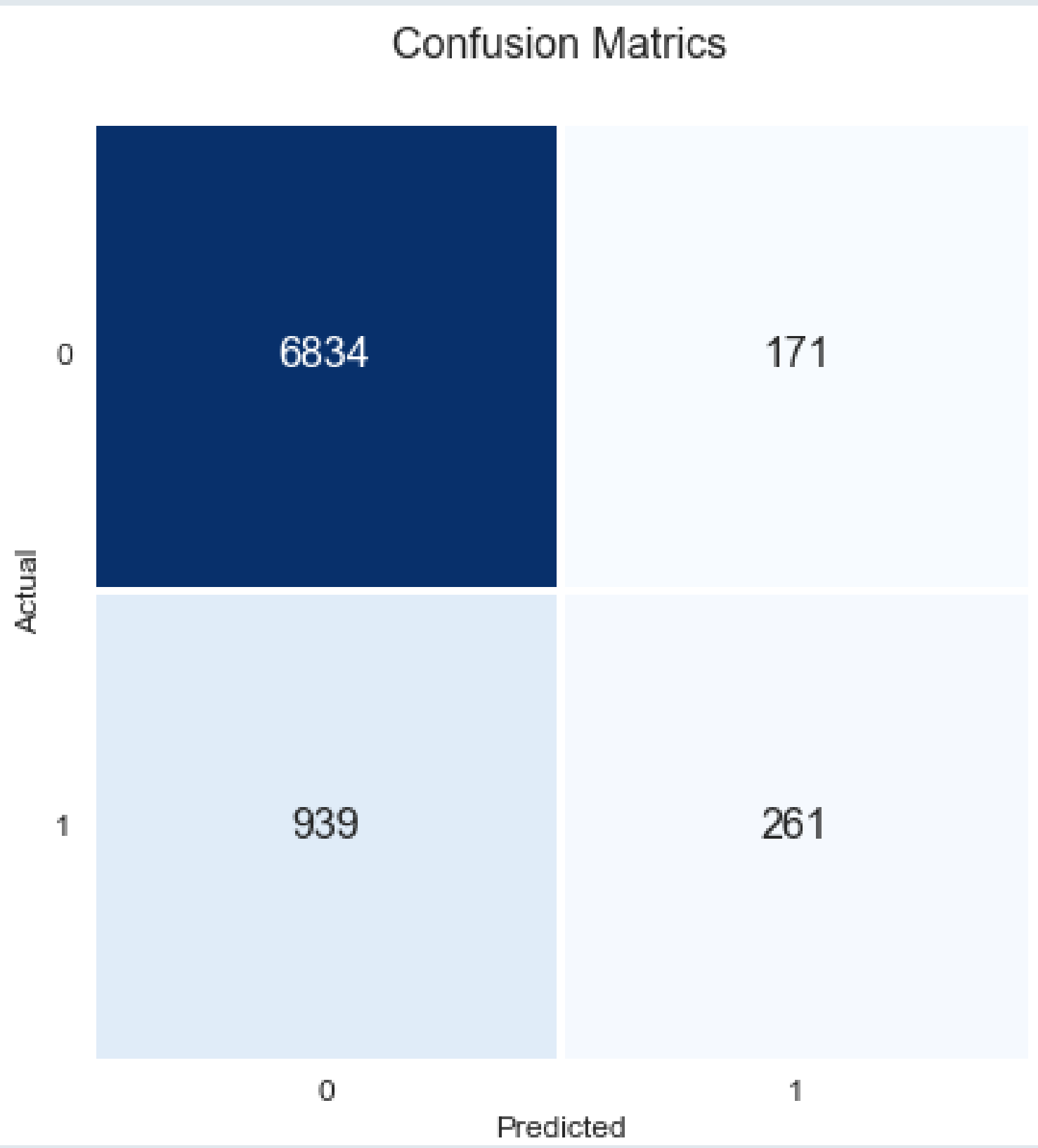
ANALIZA PODATAKA

U testnim podacima, obeležje *"Rain Tomorrow"*, za koje se vrši predikcija, u **77.6%** slučajeva ima vrednost *"No"*, dok u 22.4% ima vrednost *"Yes"*. To nam govori da su podaci uravnoteženi.



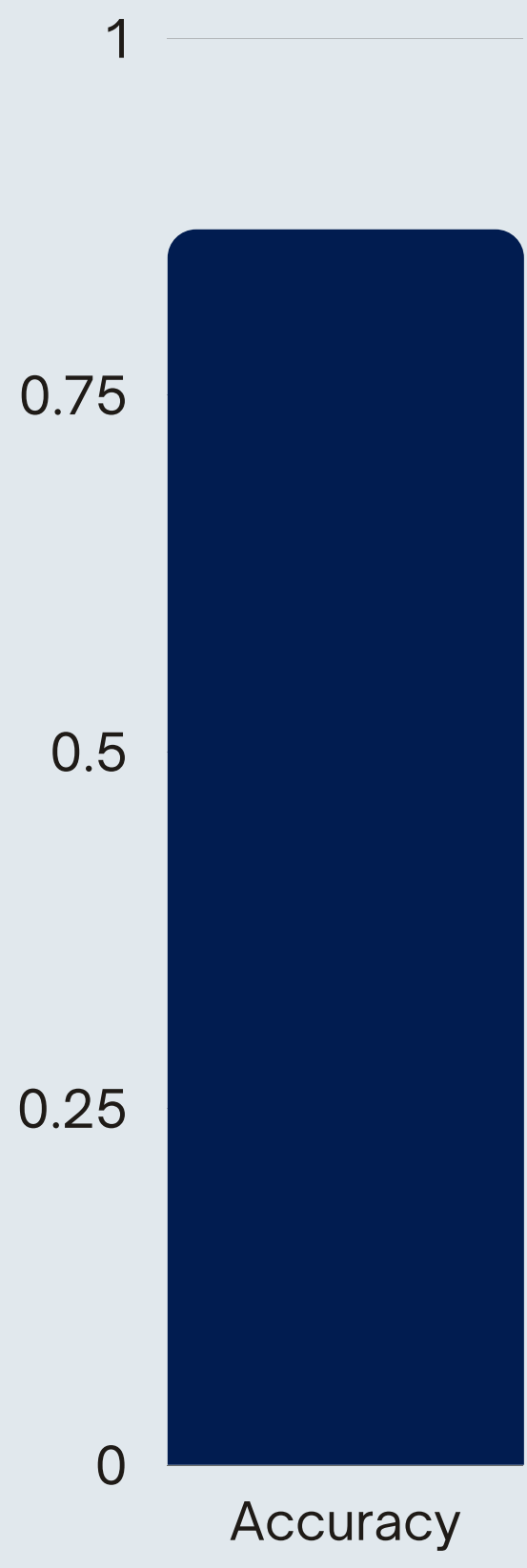
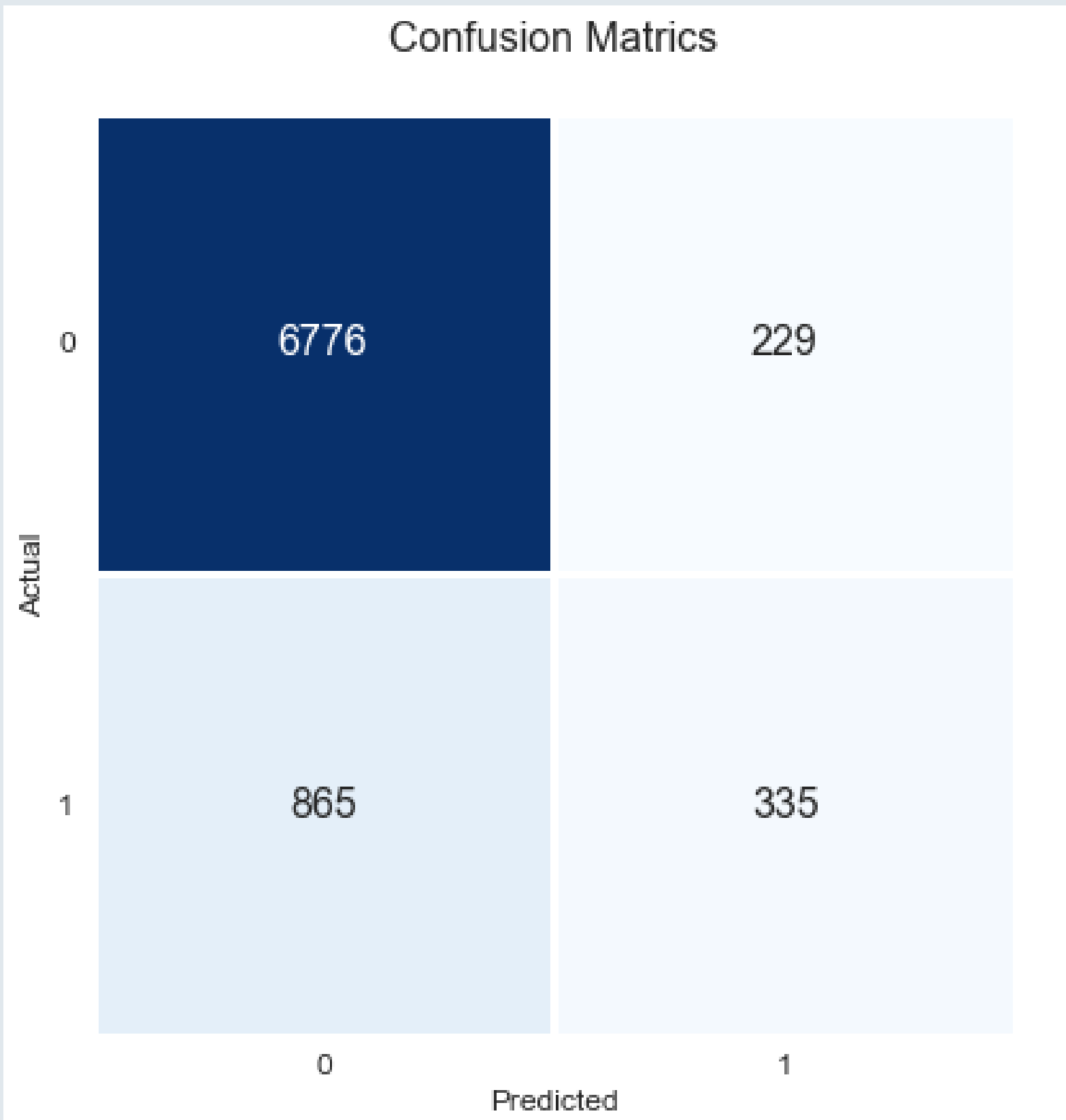
- 3267 redova ne sadrži obeležje *"Rain Tomorrow"* i oni su izbačeni
- Izbačena su obeležja koja imaju veliku međusobnu zavisnost
- Nedostajuće vrednosti su popunjene pomoću srednje vrednosti, medijane ili modusa.
- Izbačeni su outlier-i

LOGISTIČKA REGRESIJA



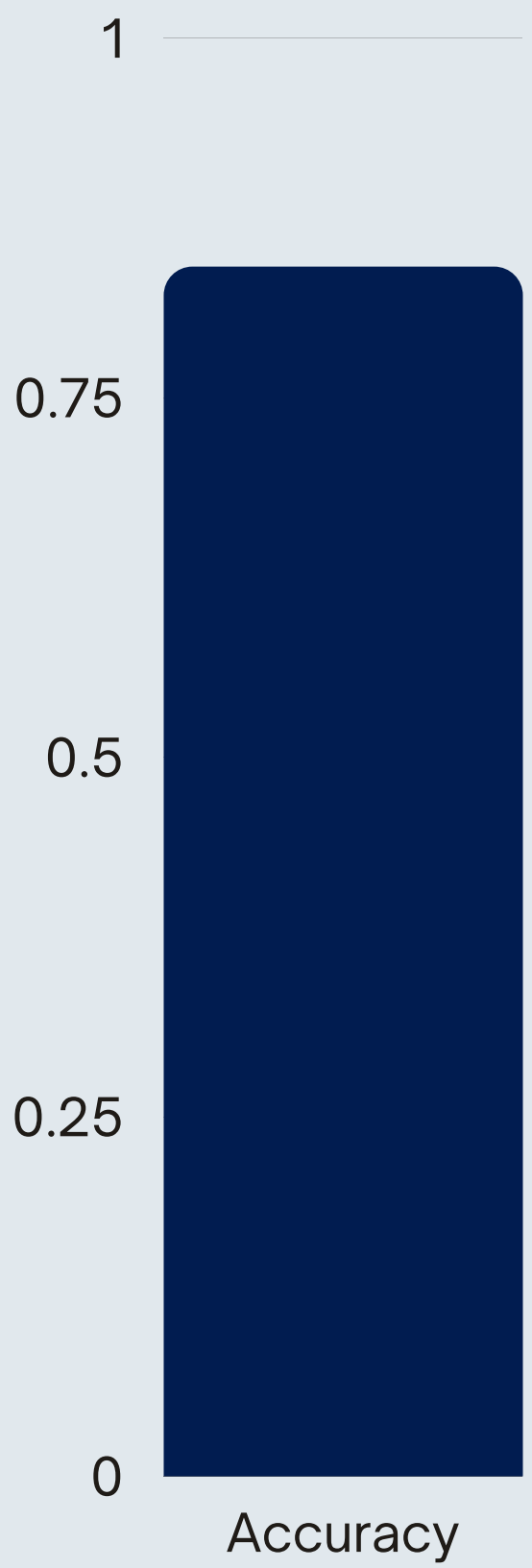
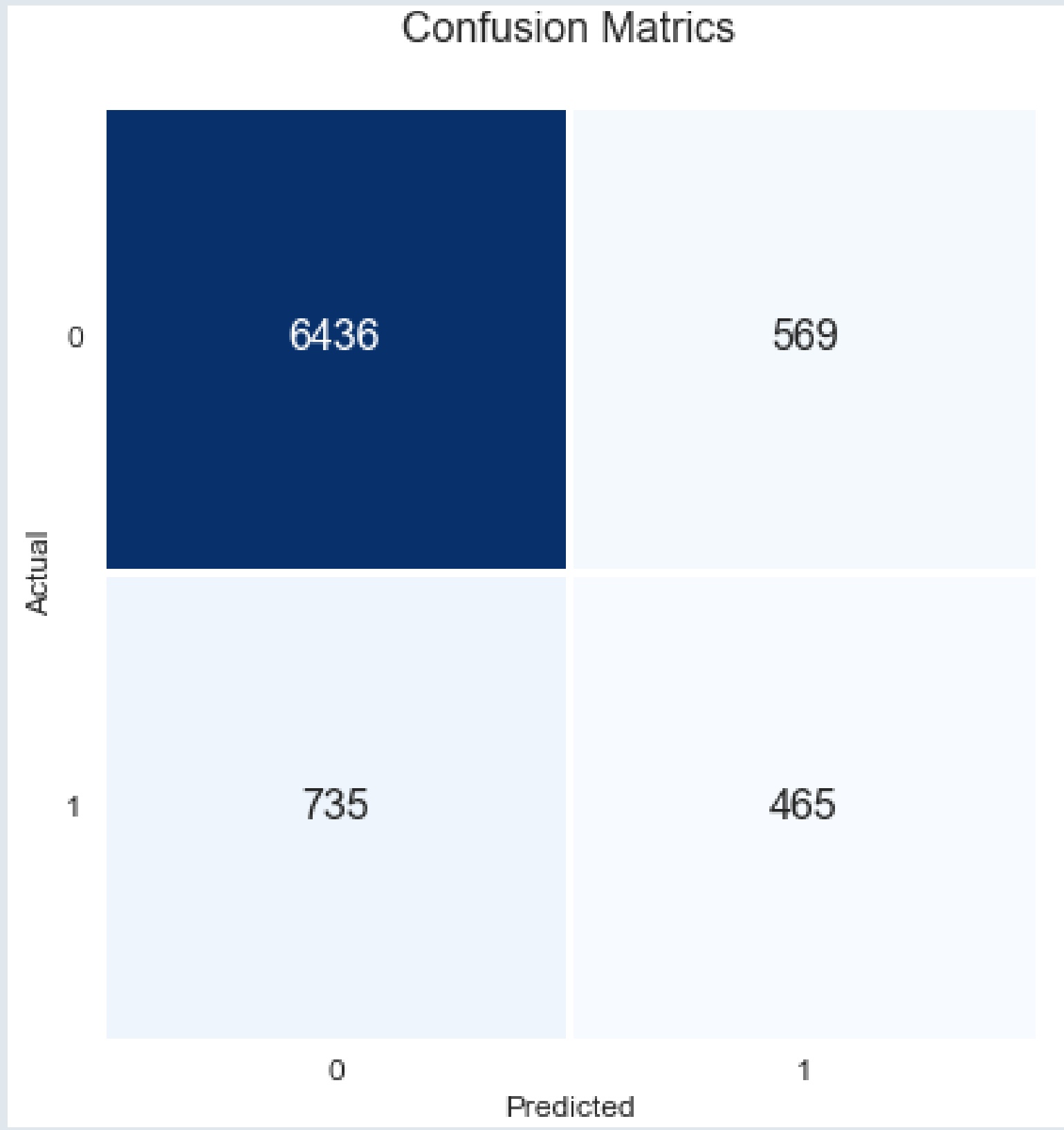
Accuracy: 86.4%

XGBOOST



Accuracy: 86.6%

ANN



Accuracy: 84.1%